

DATA MANAGEMENT PERSPECTIVE FROM MICROBOONE

MIKE KIRBY

CAVEAT EMPTOR

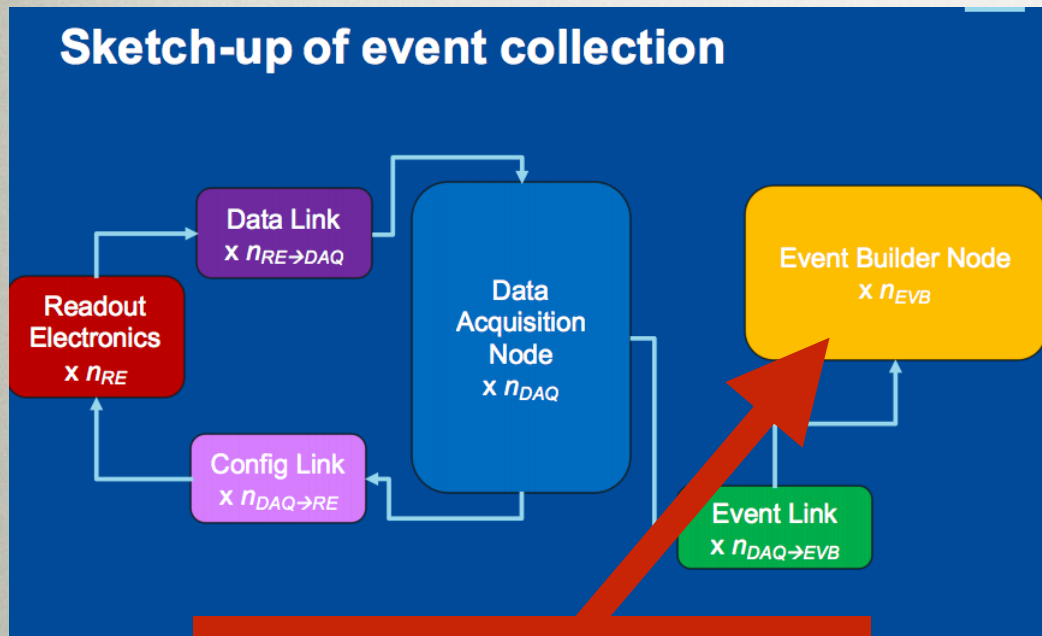
- These opinions are just from me...

DATA MANAGEMENT RESPONSIBILITIES

- data storage
- data catalog
- data integrity
- data quality monitoring
- etc.
- basically all the things you want to do once data hits disk
- understand that you're limited by hardware
- DAQ isn't the place to do everything

DATA MANAGEMENT RESPONSIBILITIES

Sketch-up of event collection

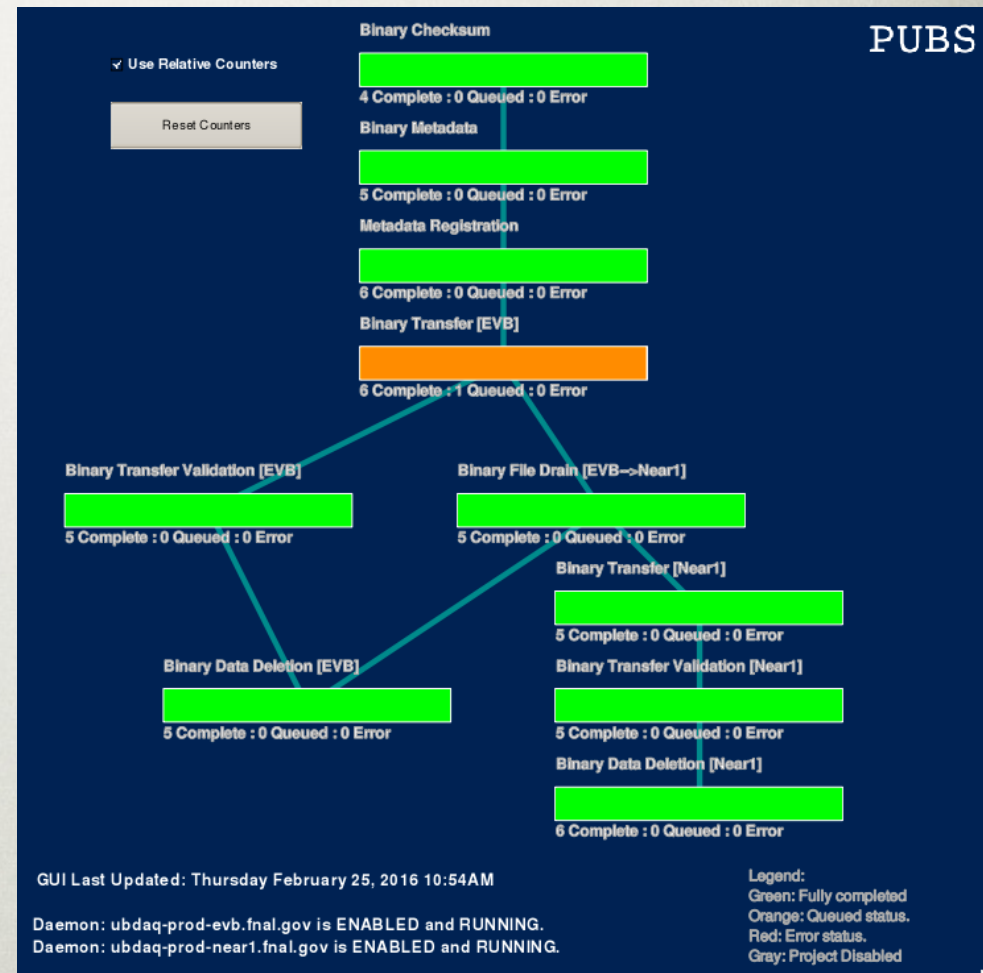


This is where DM takes over

- basically all the things you want to do once data hits disk
- understand that you're limited by hardware
- DAQ isn't the place to do everything

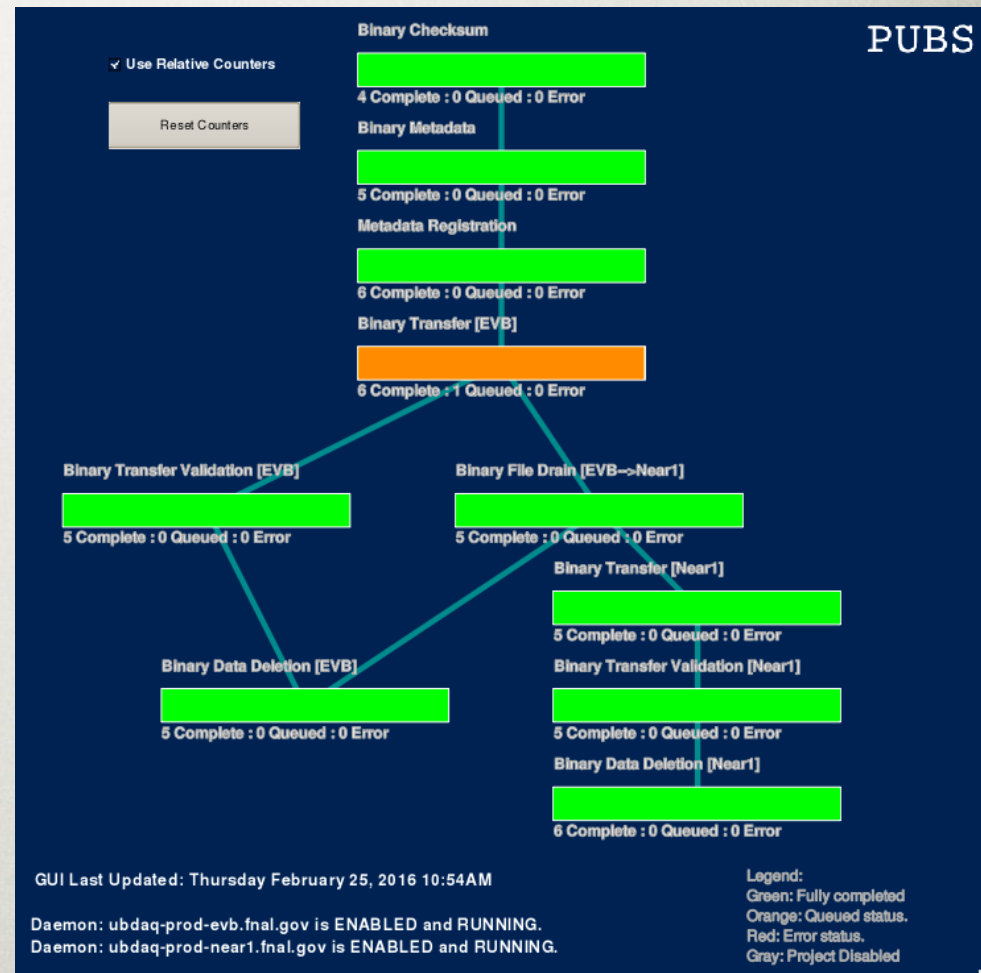
CURRENT ONLINE TASK FOR MICROBOONE

- checksum/validation
- metadata generation
- transfer files
- bypass mechanism for downtimes



CURRENT ONLINE TASK FOR MICROBOONE

- PUBS
 - crontab on steroids
 - Directed Acyclic Graphs provide dependencies
 - database backend
 - tracks all file status
 - refresh times
 - python scripts do actual work
 - limited support and/or development - Kazu is but one person...



DATA STORAGE

- permanent storage for MicroBooNE is based upon three components
 - File Transfer Service
 - SAM
 - dCache / Enstore

FILE TRANSFER SERVICE

- extremely robust transfer system
- dropbox that sorts files to the correct enstore location
- easy to develop rules and directories
- no lost files in 2.7 PB of data

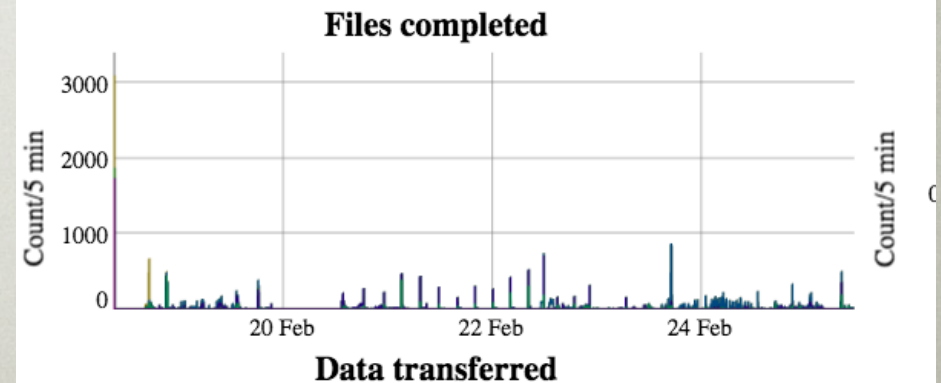
FTS status for uboone-fts-uboonesamgp

Generated at 2016-02-25 11:28:26 CST ([update](#))

Summary

FTS: OK SAM: OK

Completed files:	40330
Failed transfers:	97
All error files:	97
Waiting on tape:	532
Other pending files:	0
New files:	0



SAM

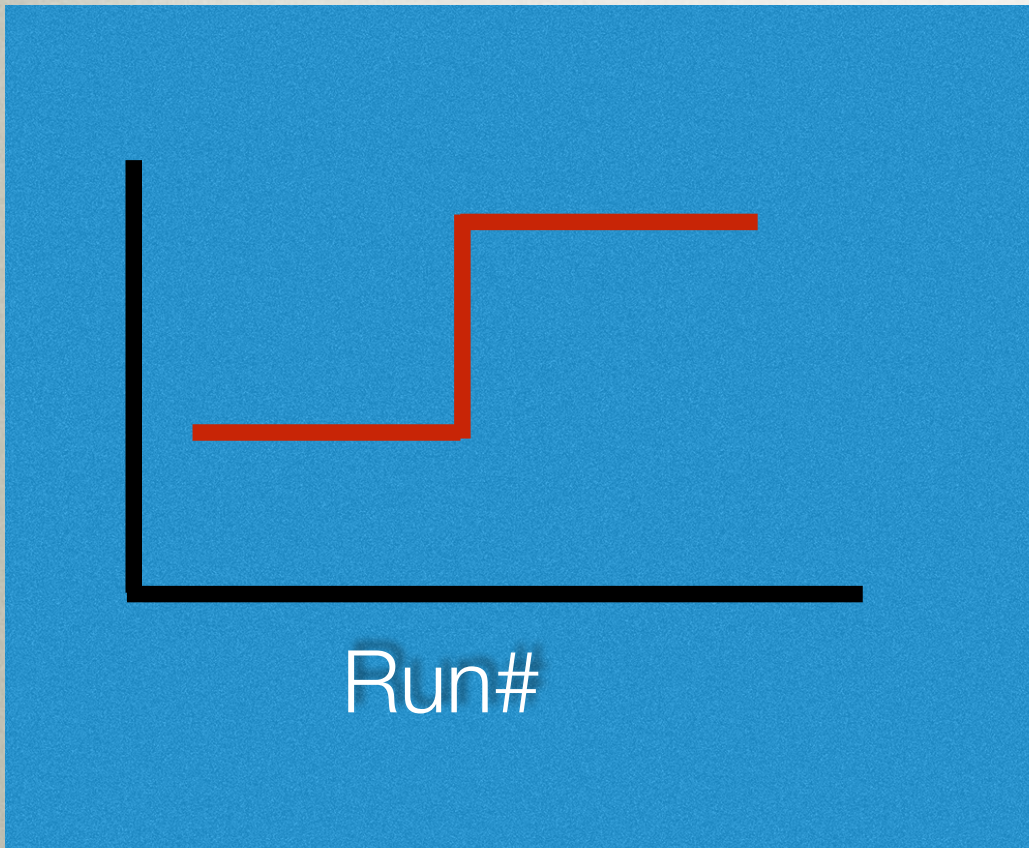
SERIAL ACCESS VIA METADATA

- this is one of many file catalog solutions, but is the backend for FTS
- can link multiple catalogs
- not using this would be rough
- Enstore even more difficult to choose otherwise

OFFLINE PROCESSING

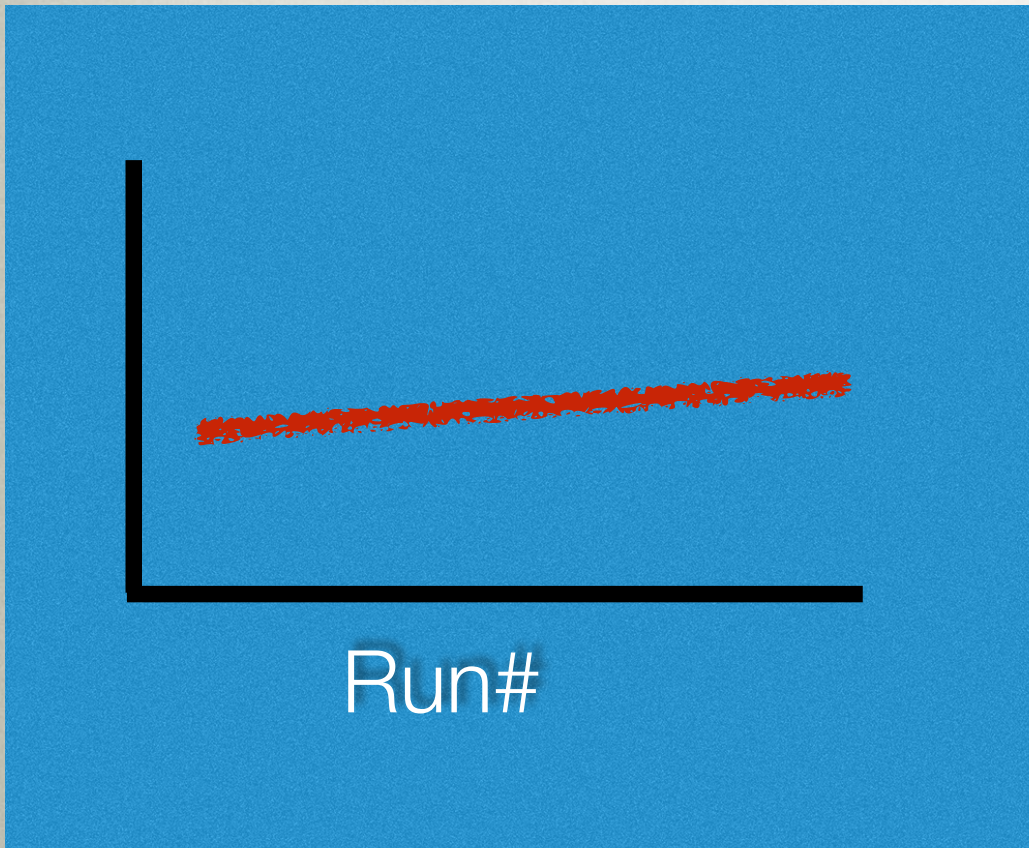
- swizzling
- reconstruction
- distribution
- cataloging output
- Data Quality Monitoring
 - two general types of DQM
 - immediate detector conditions (changes during one shift)
 - long term (changes during a week)

OFFLINE PROCESSING



- Data Quality Monitoring
 - two general types of DQM
 - immediate detector conditions (changes during one shift)
 - long term (changes during a week)

OFFLINE PROCESSING



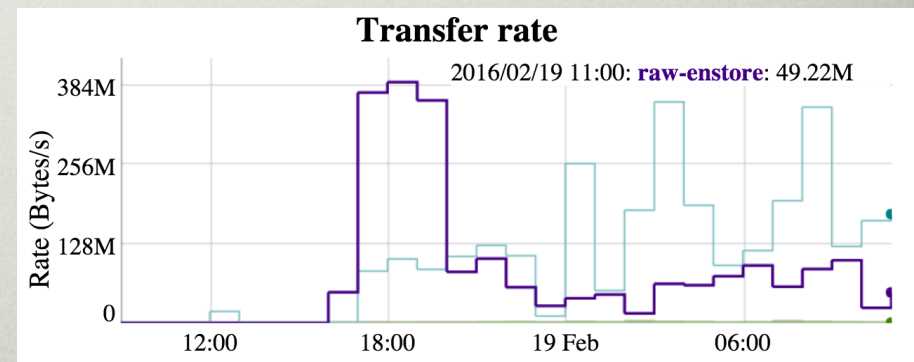
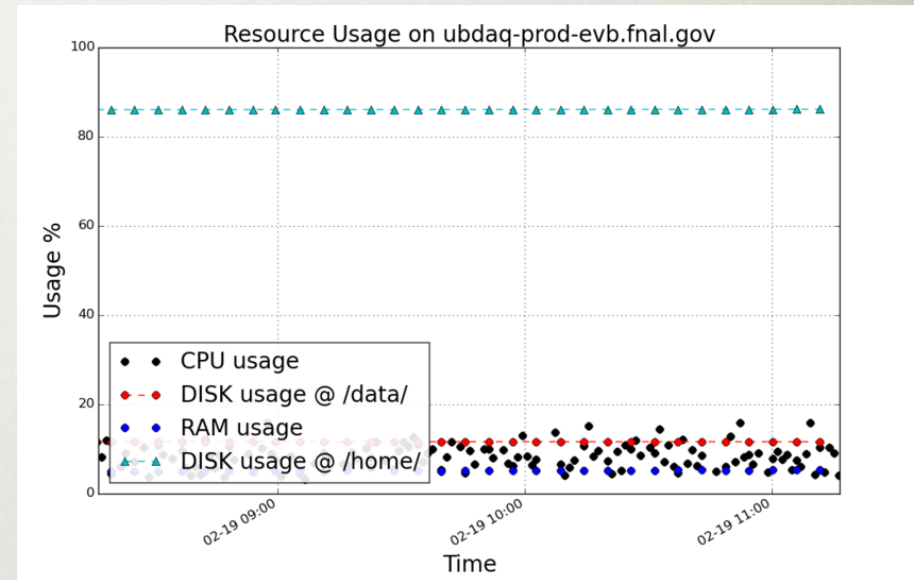
- Data Quality Monitoring
 - two general types of DQM
 - immediate detector conditions (changes during one shift)
 - long term (changes during a week)

MICROBOONE DQM

- for the nearline information utilize the artdaq online shared-memory mechanism to pick out fraction of events
- for long-term information, utilize scripts run over SAM dataset definitions

ONLINE PUBS

- running smoothly with tight trigger
- working with DAQ group to compress and store log files from /home
- successfully reconfigured for dCache downtime yesterday
- no backlog of files online
- all binary files transferred off of /data

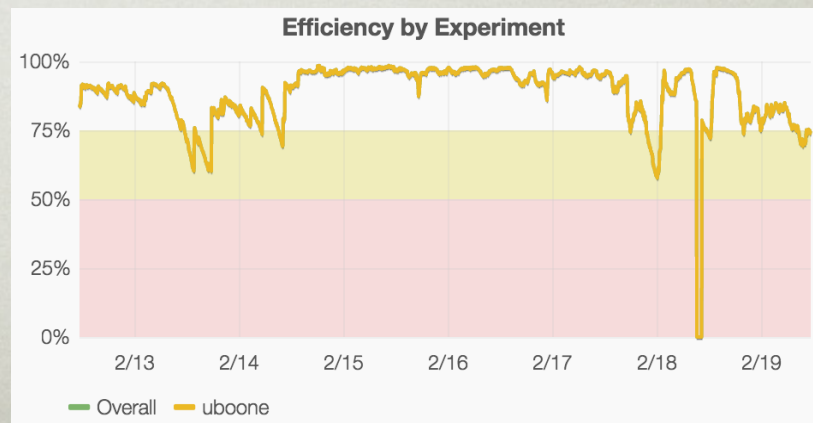
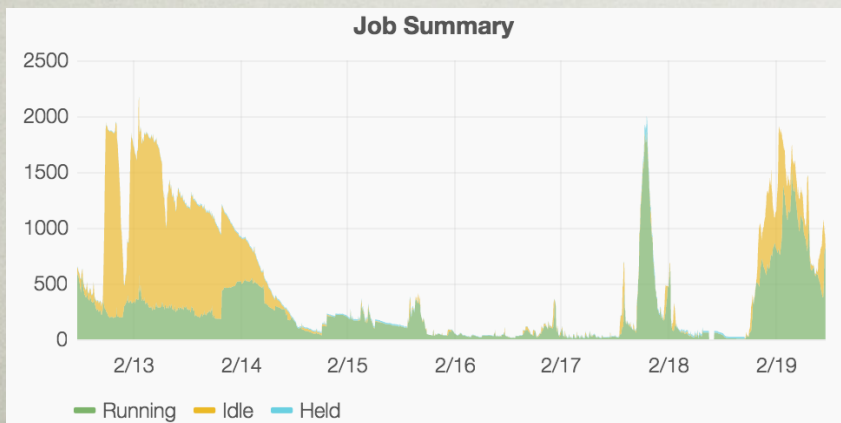


OFFLINE PROCESSING

- swizzling - v04_26_04_05 working nicely
- keeping pace with tight trigger and even caught up - good efficiency as well - last nights processing
- currently producing 7 streams - notpc, bnb, numi, extbnb, extnumi, mucs, extunbiased

Data Summary

'Stage'	'Num Evts'	'Num Files'	'Size Files(GB)'	'Tot Num Files'	'Tot Num Evts'	'Tot Size (GB)'
'Raw'	33046	662	1161	1127583	53200938	1912542
'Swizzled'	66176	931	1293	21561	9443608	31458
'Reco'	0	0	0	11652	365650	40613



OFFLINE PROCESSING

- Reconstruction processing -
v04_36_00_01 began last night after verification
- will validate outputs today / this weekend
- start producing AnaTrees next week
- working with Rui to have LArLite generation and metadata integrated

CURRENTLY PRODUCED DATASETS

- Current processing is declared v3 - datasets available here:
- http://www-microboone.fnal.gov/at_work/AnalysisTools/data/ub_datasets_v3.html
- these are the triggered samples and so our highest priority

Swizzled	Reconstructed
prod_bnb_swizzle_inclusive_v3 (describe , summary , files)	prod_bnb_reco_inclusive_v3 (describe , summary , files)
prod_numi_swizzle_inclusive_v3 (describe , summary , files)	prod_numi_reco_2016week2 (describe , summary , files)
prod_extbnb_swizzle_inclusive_v3 (describe , summary , files)	prod_extbnb_reco_inclusive_v3 (describe , summary , files)
prod_extnumi_swizzle_inclusive_v3 (describe , summary , files)	prod_extnumi_reco_inclusive_v3 (describe , summary , files)
prod_mucs_swizzle_inclusive_v3 (describe , summary , files)	prod_mucs_reco_inclusive_v3 (describe , summary , files)
prod_extunbiased_swizzle_inclusive_v3 (describe , summary , files)	prod_extunbiased_reco_inclusive_v3 (describe , summary , files)

Libo Jiang

OLDER DATASETS

- All v2 datasets (Jan 1 - Feb 11) http://www-microboone.fnal.gov/at_work/AnalysisTools/data/ub_datasets.html
- Good Run datasets (Jan 1 - Feb 11) http://www-microboone.fnal.gov/at_work/AnalysisTools/data/ub_data_v2_goodrun.html

GOOD RUN TABLE PAGE

Run	Run Type	Good For BNB Analysis	Taking BNB Triggers	Avg BNB Intensity	Avg BNB Horn Current	Avg BNB Trigger Rate	Good for NuMI Analysis	Taking NuMI Triggers	Avg NuMI Intensity	Avg NuMI LINA Current	Avg NuMI LINB Current	Avg NuMI LINC Current	Avg NuMI LIND Current	Avg NuMI Trigger Rate	PMT HV On	Wire Bias On	Avg Drift HV	Min Drift HV	Max Drift HV	Start	End
5000	Physics	Yes	Yes	4.66E12	172.8 kA	3.47 Hz	Yes	Yes	38.4E12	-49.5 kA	-49.9 kA	-49.8 kA	-49.9 kA	0.65 Hz	Yes	Yes	69.87 kV	69.79 kV	69.94 kV	2016-02-14 05:35:45	2016-02-14 09:43:35
5001	Physics	Yes	Yes	4.69E12	174.1 kA	4.23 Hz	Yes	Yes	38.3E12	-49.5 kA	-49.9 kA	-49.8 kA	-49.9 kA	0.66 Hz	Yes	Yes	69.87 kV	69.79 kV	69.94 kV	2016-02-14 09:46:05	2016-02-14 12:32:17
5002	Physics	Yes	Yes	4.69E12	174.3 kA	4.23 Hz	Yes	Yes	38.7E12	-49.5 kA	-49.9 kA	-49.8 kA	-49.9 kA	0.65 Hz	Yes	Yes	69.87 kV	69.79 kV	69.94 kV	2016-02-14 12:34:46	2016-02-14 14:23:29
5003	Physics	Yes	Yes	4.65E12	173.0 kA	4.21 Hz	Yes	Yes	37.0E12	-49.5 kA	-49.9 kA	-49.8 kA	-49.9 kA	0.65 Hz	Yes	Yes	69.87 kV	69.79 kV	69.94 kV	2016-02-14 14:26:19	2016-02-14 14:36:02
5004	Physics	Yes	Yes	4.69E12	174.3 kA	3.90 Hz	Yes	Yes	38.7E12	-49.5 kA	-49.9 kA	-49.8 kA	-49.9 kA	0.56 Hz	Yes	Yes	69.87 kV	69.79 kV	69.94 kV	2016-02-14 14:38:54	2016-02-14 14:50:07
5005	Physics	Yes	Yes	4.65E12	173.3 kA	4.23 Hz	Yes	Yes	38.2E12	-49.5 kA	-49.9 kA	-49.8 kA	-49.9 kA	0.65 Hz	Yes	Yes	69.87 kV	69.79 kV	69.94 kV	2016-02-14 14:52:51	2016-02-14 21:52:57
5006	Physics	Yes	Yes	4.69E12	174.3 kA	3.80 Hz	Yes	Yes	38.7E12	-49.5 kA	-49.9 kA	-49.8 kA	-49.9 kA	0.55 Hz	Yes	Yes	69.87 kV	69.79 kV	69.94 kV	2016-02-14 21:55:31	2016-02-14 21:57:06
5009	Physics	Yes	Yes	4.67E12	174.0 kA	4.14 Hz	Yes	Yes	38.3E12	-49.5 kA	-49.9 kA	-49.8 kA	-49.9 kA	0.65 Hz	Yes	Yes	69.87 kV	69.79 kV	69.94 kV	2016-02-14 22:04:06	2016-02-14 22:32:39
5010	Physics	Yes	Yes	4.68E12	174.1 kA	4.27 Hz	Yes	Yes	38.4E12	-49.5 kA	-49.9 kA	-49.8 kA	-49.9 kA	0.65 Hz	Yes	Yes	69.87 kV	69.79 kV	69.94 kV	2016-02-14 22:36:13	2016-02-14 23:30:53
5011	Physics	Yes	Yes	3.27E12	121.5 kA	4.00 Hz	Yes	Yes	27.4E12	-49.5 kA	-49.9 kA	-49.8 kA	-49.9 kA	0.60 Hz	Yes	Yes	69.87 kV	69.79 kV	69.94 kV	2016-02-14 23:34:03	2016-02-15 01:06:36
5012	Physics	Yes	Yes	0.11E12	4.3 kA	1.22 Hz	No	Yes	0.0E12	0.0 kA	0.0 kA	0.0 kA	0.0 kA	0.27 Hz	Yes	Yes	69.87 kV	69.79 kV	69.94 kV	2016-02-15 01:10:23	2016-02-15 01:42:29
5013	Physics	Yes	Yes	4.62E12	169.6 kA	4.18 Hz	Yes	Yes	38.0E12	-49.5 kA	-50.0 kA	-49.8 kA	-50.0 kA	0.66 Hz	Yes	Yes	69.87 kV	69.79 kV	69.94 kV	2016-02-15 01:45:54	2016-02-15 08:45:56

Nick Graf

MONTÉ CARLO PRODUCTION

- MCC7 is in full swing - focusing on high priority samples: BNB, BNB+cosmic, cosmic, BNB intr nu_e
- samples are here: http://www-microboone.fnal.gov/at_work/AnalysisTools/mc/mcc7.0/
- working on processing as much as possible offsite

MC Summary 'Stage'	'Num Evts'	'Num Files'	'Size Files(GB)'	'Tot Num Files'	'Tot Num Evts'	'Tot Size (GB)'
'Gen'	0	0	0	20	200	0
'Sim'	0	0	0	20	200	0
'DetSim'	0	0	0	20	200	9
'Reco'	20760	1038	1914	17183	607300	61671

Joel Mousseau